A Hybrid Weighted Ensemble classifier model for medical databases

P R Sudha Rani[1], Dr. K R Ramya[2]
[1]Research Scholar, [2]Professor
[1,2] Department of CSE, Acharya Nagarjuna University (University College of Engineering and Technology) Nagarjuna Nagar-522510, Guntur, A.P, India
[1]Email: sudharani.p@gmail.com

ABSTRACT

Extreme learning approaches are now widely used to identify and diagnose medical conditions for large databases. Ensemble classifier was a key study model for extreme learning machines for real-time applications due to its great performance and processing speed. Due to the static weight selection of the output layer hidden, standard extreme learning methods are unable to estimate the error rate. This research introduces a novel weighted extreme learning machine (WELM) for medical condition prediction. The basic goal of the weighted extreme learner is to define high-dimensional data for illness prediction. Typically, the proposed ensemble model is created and deployed to improve cancer prediction using high-dimensional data. Using several ensemble learning models such as random forest, neural networks, ACO+NN, and PSO+NN, we evaluated the performance of the WELM model suggested in this paper. Test outcomes are examined in a variety of medical datasets, including liver, diabetes, ovarian, and DLBCL-Stanford. The WELM presented is highly computationally efficient in terms of true positive rate, error rate, and accuracy, according to the results.

Keywords: Extreme learning machine, Neural network, Ensemble classifier.

1. Introduction

A feed-forward neural network is the most extensively used and best recognised neural network. A hidden layer or layers and an output layer are present in the system. The output layer[1] transmits the final response to the training dataset. The topic of neural feed-forward networks has seen a lot of activity in recent years. The model's inputs directly characterise its dynamic linear or nonlinear structure. The basic parametric restrictions of such structures make them unsuitable for dealing with big inputs in traditional models. Another important characteristic of the neural feed network is the mapping of parameters between layers.

This repeated nature of parameter adjustments in the standard SLFN technique might lead to problems. Using the Extreme Learning Machine (ELM) approach, these concerns can be addressed. Each layer's weights and hidden nodes are chosen at random in SLFN before the final output is evaluated. Consequently, the ELM technique has a superior generalisation performance and a faster learning speed.

In comparison to non-parametric approaches, most classical learning models for training SLFNs are considerably slower than those used nowadays. This method takes a long time to work because it requires a lot of fine-tuning of settings. Furthermore, these models demand a large amount of computational memory and increase the overall computational time required for the mapping process. SLFNs are still used in Extreme Learning Machine, although it's a slightly modified version of the original [2-4]. Conventional SLFNs can benefit from this technique because it improves efficiency and performance. There is also a great deal of manual tuning of control parameters (such as learning rate and epochs) and local minima in most neural network learning systems. In contrast, manual iterative adjustment is unnecessary while using Extreme Learning Machine[5]. There is a problem with the classification border in ELM and the boundary remains constant during the training phase. As a result, there is a risk of incorrectly classifying samples that are near the boundary. In comparison to other classic tuning-based techniques, this method necessitates a much higher number of hidden neurons [6]. The standard neural network paradigm for data classification has been extended by Extreme Learning. It breaks down the problem into a series of smaller issues, and then combines them to find the optimum answer for each of them. Samples of training data are stored in the parameters of the output layer, which are hidden from view. There are two types of extreme learning machines: the "Weighted ELM" and the "Technical ELM."

The weighted-ELM method was then developed to address the shortcomings of traditional ELM. For huge datasets, the weights are increased over time. A variety of learning methodologies can be used to tailor the parameters of each layer in most feed-back ANN systems. Many significant learning methods in feed-in neural networks are based on BP techniques and gradient descent[7-8]. In comparison to ANNs, neural feed networks have a much slower rate of model learning. ELM is referred to as "Extreme Learning Machine" because of its enhanced generalisation and speed of computation (ELM). Problems with standard gradient decreasing methods, such as stop criteria and learning rate [9] and local minimum [10], have been identified. These issues are addressed by the Extreme Learning Machine. Research in the areas of health and biomedicine has recently focused on the examination of cancer diagnosis data. There is currently no cure for all forms of cancer. As a result, physicians and patients alike must rely on early disease identification to help prevent the disease from progressing. When it comes to cancer diagnosis, microarray gene expression datasets include a wide range of genetic expressions that can be used. A feature of ELM based extraction and classification relies heavily on the extraction of characteristics. There are two types of characteristics: those that fall into one of two categories: The first group's extraction features rely on noisy attributes and contextual data. Secondly, the second group shares some traits with the first.

Medical data prediction has become one of the most crucial and complex criteria in recent years. There are a number of accepted methods for predicting medical disorders, such as cancer. Breast cancer risk can be predicted and detected using a technique based on association rule mining, MLP, and background propagation . The GSA algorithm and the embedded fluke logic method are utilised to detect and evaluate heart disease in a modular neural network[3]. As information technology advances, data mining approaches are being used in a variety of fields, including biomedical applications and disease prediction. Detection and early diagnosis of cancer should be given special attention. For the most part, a gene selection strategy is used in conjunction with current pattern categorization methods. Medical data sets have a lower mistake rate when the noise is reduced.

2.Related Work

Traditional techniques' over-fitting problem has been addressed by a novel strategy that integrates neural networks. In [2], a voting-based ELM technique, the overfitting problem is likewise overcome. The ultimate choice is made by majority vote after numerous independent ELM trainings have been received. The underlying principle of this strategy is to randomly select a portion of data to create a new dataset I. Afterwards, it selects a tiny subset of S's feature subspace. Using ELM and feature space subset S, we build a classifier for training dataset I.[6] created the first iteration of the ELM strategy. Conventional SLFNs learning methods have disadvantages including slower learning speed, trivial parameter adjustment and poor generalisation capacity, which this approach aims to fix. With superior generalisation performance, the approach is commonly used in the fields of classification and regression. While classic Gradient-Based Iterative Learning techniques like Back Propagation are thought to be slow, ELM is considered to be a more effective generalisation method. Extremely fast learning, excellent generalisation, and complete parameter freedom are the hallmarks of the new ELM algorithm [5].Gene expression classification and classical classification challenges can be solved using a new ensemble machine learning approach proposed in [8]. The morphological appearance of cancer-related datasets is taken into account by most classic classification methods. The classification method is carried out using a single C4.5, Bagging, and AdaBoost decision trees in this study. Clustering techniques are responsible for determining the similarity index between genes or a collection of genes under the same conditions.Text classification often contains data with an unbalanced distribution of classes. In general, it categorises a lot of garbage, although there are a few items that fall into the "interesting" category. The majority of BN techniques are built using standard classifiers. These methods produce precise results and allow for the representation of relationships among variables. The typical class-imbalanced problem cannot be solved by this method. Many innovative approaches have been developed to improve the accuracy of standard classification techniques in order to solve the problem of class imbalance. Sample approaches, cost-sensitive techniques, recognition-based method and active learning technique are some of the generalised classification approaches. Sampling methods have been developed to address class imbalance issues by excluding specific data from the dominant class. Under sampling is another name for this method. An over-sampling issue occurs when a few additional artificially generated data points are included in the minority class. In general, cost-matrix for all types of errors or instances is used in cost-sensitive learning methods. Its primary goal is to facilitate mechanisms for learning from unbalanced data sets. Oversampling the minority class has the same impact as the aforementioned mechanism. Overly precise regulations or rules that are overly tailored to training can be the result. Random under sampling is a method of removing a random number of students from an overly large class. The aforementioned procedure is repeated until the class size and cost-consciousness of the other class is met. It incorporates the changes in the relative cost associated with misclassification of positive and negative classes. The results of both strategies are evaluated and compared to the results attained without balancing.Rules from the minority class can be learned using methods based on recognition.

The majority class may or may not be used in this example. Unlabeled training data necessitates the use of active learning strategies to help students learn more effectively. ELM's generalizability is tested by optimising output weights for the minimum possible training error and output weight [5]. ELM's universal approximation capability is always met. Other applications include system modelling, prediction, control, robotics and computer vision as well as medical and biomedical sciences. ELM optimization is also used in these various fields. The Fuzzy ELM (F-ELM) approach consists of two stages: preparation and prediction. During training, a three-tiered architecture is presented. Iteratively generating hidden layer, hidden layer, and output layer scaling parameters and weights is a common practise in model-based learning (ELM). The functions of F-ELM in regard to fuzzy inference are also identical. F-ELM algorithms are used to map feature inputs into an F-ELM algorithm in order to produce predictions.For the binary categorization of microarray datasets, a new ELM-based technique was proposed[11]. All/AML, CNS, lung, ovarian, and prostate cancer datasets were analysed in this study. [http://www.eps.upo.es/big/dataset.html] Binary classification performance has been analysed. The first step was to do feature extraction using the correlation coefficient. ELM has been proven to be a better choice than many other standard classification algorithms, according to the researchers. First and last datasets have slightly lower classification accuracy, however classification accuracy for the other three datasets is higher. There is a weak association between these two datasets, which is why their precision is lower.Differential Evolution (DE) Method is an enhanced version of Extreme Learning Machine (ELM), a form of global optimization algorithm. [8] Evolutionary ELM was presented as a modified version of the classic ELM technique (E-ELM). The Differential Evolution method is used in this approach to find the best network parameters. They designed a self-adaptive Evolutionary Extreme Learning Machine (EELM) to enhance the basic ELM (SaE-ELM). This method is in charge of identifying the best control settings and DE generation methods. [10] Traditional ELM was reimagined as developing ELM in a creative way (Evo-ELM). Control parameters can be fine-tuned using this strategy. Increasing the number of predictors stabilises and improves the accuracy of this method. Tuned parameters are required in every classic Differential evaluation technique. It is not necessary to tune parameters in evolutionary learning models since the parameters of hidden layers are randomly selected. As opposed to alternative techniques, such as quadratic programming in SVM or gradient descent in Back Propagation-based Neural Networks, regularised least squares solves problems faster.

To create an ELM scheme that is as efficient as possible, numerous theoretical and practical studies have been offered. [9] Both semi-supervised and unsupervised learning approaches were used to implement the ELM technique. The manifold regularisation is used to implement this strategy. The ELM technique to clustering takes into account unlabeled and partially-labeled samples in this case.[5-8] Using kernel learning and typical optimization techniques, we were able to enhance the traditional ELM strategy. Like SVM but with higher scalability, lower computational complexity, and optimization restrictions this technique is superior than the SVM model In addition, this method is unable to accomplish any specific goal. When it comes to predicting the lymph node status of breast cancer patients, genetic algorithms and neural networks are combined [13]. ELM was used to identify the location of breast tumours in digital mammography datasets. Wavelet transformation is used to classify micro-classified mammograms as well.The feature space ensemble has been added to the ELM model in this study. This is a more sophisticated method known as the weighted ELM based ensemble extreme learning machine (W-EELM). This method's major goal is to integrate weak classifiers into a strong classification scheme in order to improve its performance.

3. PROPOSED MODEL

In this paper, a novel weighted extreme learner was developed for ensemble learning classifier to address the problem of true positive rate and accuracy issues. The main objective of the weighted extreme learner is to classify the high dimensional data for disease prediction.

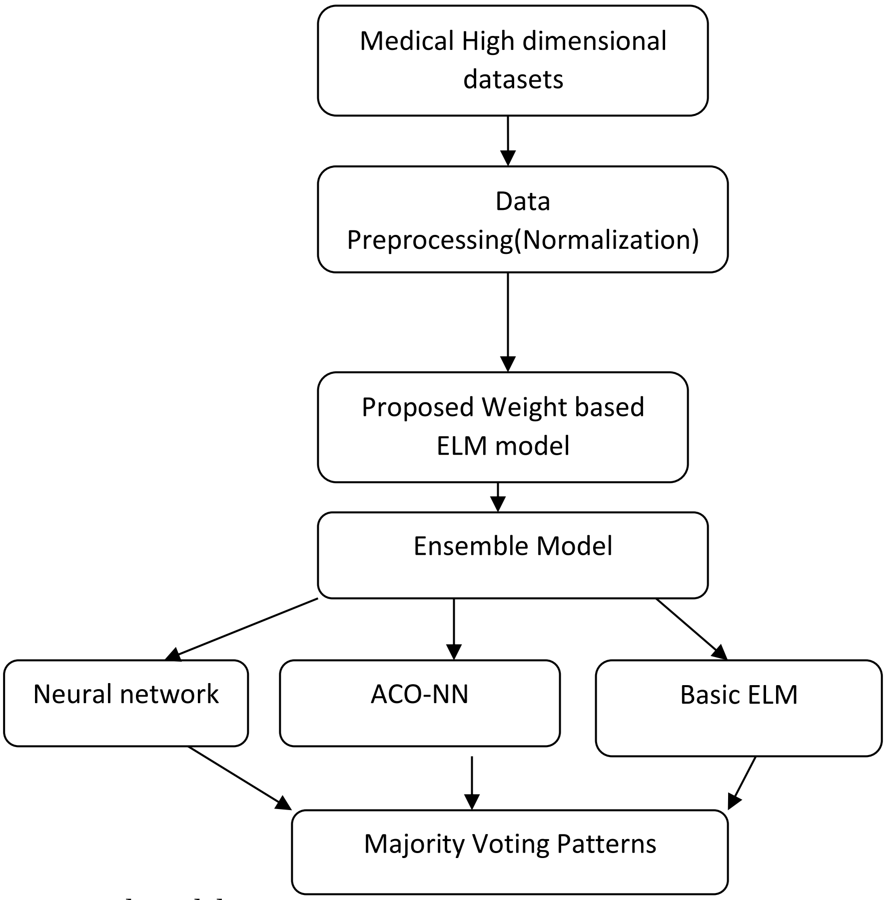

Figure 1: Proposed Model

For high-dimensional data, the proposed ensemble model is often built and deployed to increase the accuracy of the cancer prediction rate. The proposed ensemble model enhances the accuracy and true positive rate of the classic ensemble classifiers; NN, ACO-NN and ELM classifiers.

Figure 1 depicts the suggested model's basic workflow. This model compares the proposed model's performance against that of the current models using a variety of high-dimensional datasets. To begin, the input datasets are normalised and the sparsity problem is addressed. The

suggested weighted ELM model is used as an ensemble classifier on the training data after normalisation. The classic classifiers Back Propagation (BP)NN, ACO-NN, and ELMs are improved with the proposed weighted ELM model for the ensemble classifier.
The following characteristics have received particular attention in this study:-

1) The suggested method's learning speed is superior to standard methods like Back Propagation (BP)NN, ACO-NN, and ELMs.

For high dimensional datasets, this model can be applied because typical ELM techniques are sensitive.

Instead of dealing with an enormous number of hidden nodes, our proposed strategy with m hidden nodes can improve training accuracy.

There is evidence that using only m hidden nodes can produce the same or greater generalisation performance as standard techniques that use many hidden nodes.

Input : Input Data T_r.

Output: Preprocessed Data.

Procedure:

Initialize training dataset $T_{r.}$
For each attribute A(i=1,,n) in T_r
Do

If(A[i] is numeric)

Then

For each attribute value in A[i]

do

If(A[j]!=null)

$$SNormalize(A[j]) = |A[j] - \mu_{A[j]}| / (Max\{\mu_{A[j],S[m]}\} - Min\{\mu_{A[j],S[m]}\})$$

Else

$$SNormalize(A[j]) = |A[k] - \mu_{A[k]}| / (Max\{\mu_{A[k],S[m]}\} - Min\{\mu_{A[k],S[m]}\}); \text{ where } j \neq k$$

End if

If(A[j] is nominal)

Then

$$SNormalize(A[j]) = \sum Prob(A[j] / S_m) / (Max\{\mu_{A[j],S[m]}\} - Min\{\mu_{A[j],S[m]}\})$$

Where S_m represents the m –classes

$\mu_{A[j],S[m]}$ Represents the mean of the each mth class.

End if

Done

Algorithm2: Proposed Weighted ELM algorithm based Ensemble Classifier

In its simplest form, an extreme learning machine is a feed forward network with a single hidden layer that uses standard data to establish instance weights. Instance biases on the linked items are computed during weight initialization. Error-based updates are at the heart of weights and biases. Intense learning machines have a large impact on the evaluation of the model based on the number of neurons. If both the number of neurons and the error rate are high, the suggested model will reduce the training data set's error rate.

Input: Normalized data ND, Max iterations Max, bias λ, Weighted vector W, c is class label ,I is instance , I[c] is instance class, I[t]: t Attribute instance value.

Output : class label prediction.

Procedure:

Initialize weights to all instances.

For each instance I in D do

$R \leftarrow \sum_{d=1}^{\#Attributes} \log(W_d).I_t + \lambda$

Done

if R(log(W)I+bias)<0

Then

For each attribute t do

// update previous cached weights

$W_{p,t} \leftarrow |\log(W_{p,t})| + R.C_t.I_t$

// update cached bias values

$\lambda_p = \lambda_p + I[c].C_t$

End if

N_m^k []=Find Top K- nearest objects from Sorted list of Dist;

Assign a class to p' based on majority vote: $c'=argmaxy\sum(xi,ci)$ belonging to S, $I(y=ci)$

Done

return $(W-\frac{W_{p,t}}{\sqrt{C_t}}, \lambda - \frac{\lambda_p}{\sqrt{C_t}})$

Where, R' always greater than 1.

Algorithm 3: Ensemble Steps

$D_0 \leftarrow <1/D_1, 1/D_2, 1/D_3, \ldots 1/D_N >$ N uniform data partitions
$C_1 \leftarrow$ Proposed Algorithm
$C_2 \leftarrow$ ELM
$C_3 \leftarrow$ ACO – NN
$C_4 \leftarrow$ Neural network

for each classifier k in C_i do
$C_i = \text{classifier}(D, D_{k-1})$ // Training phase

$\hat{P}_n \leftarrow C_i(x_n)$, $\forall n$ prediction on training instances

$\hat{E}_k \leftarrow \sum_n D_{k-1}[n] \ / \ P_n \neq \hat{P}_n$

$\phi_k = \log\left(\frac{1-\hat{E}_k}{\hat{E}_k}\right)$

end for

return $f(\hat{m}) = \text{sgn}[\sum \phi_k . C_k(\hat{m})]$

4.Experimental Results

Using the proposed ensemble model, experimental findings are shown below. A list of datasets used in the experiment may be seen in Table 1, which provides an overview[12]. For cross-validation, 10% of the training data is utilised as a test dataset. There are a variety of data sets utilised in the experiments to ensure that the proposed model's interpretation is accurate.

Table 1: Datasets and its properties

Dataset Name	Features	Type
lung Cancer	12533	Continuous
DLBCL-ovarian	4219	Continuous
lung-Michigan	7230	Continuous

Sample Ovarian Dataset

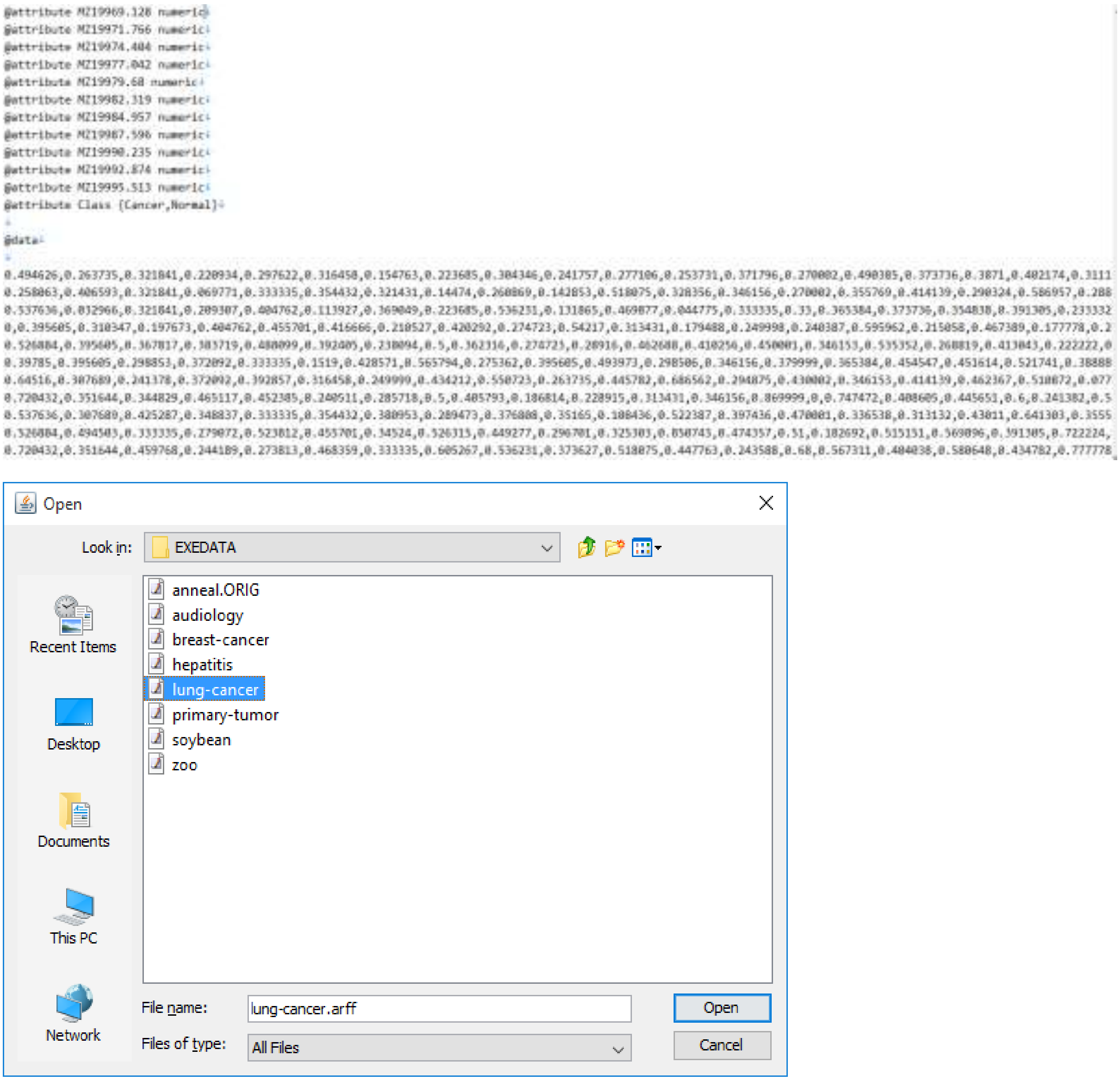

Figure 2: Input dataset

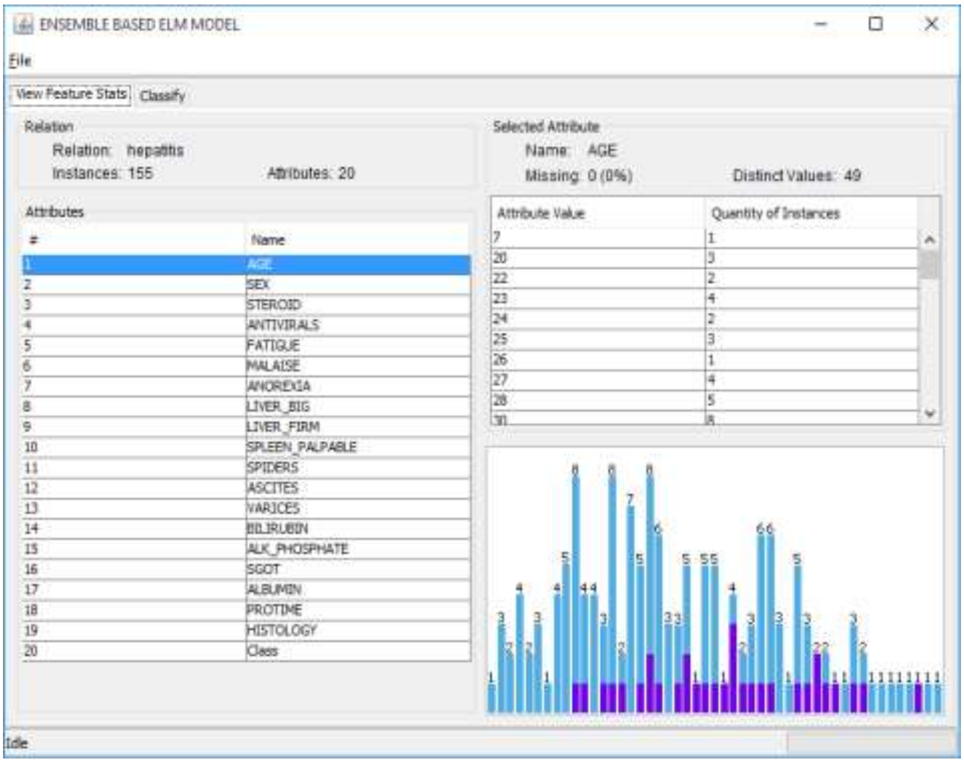

Figure 3: Statistical analysis of the input dataset.

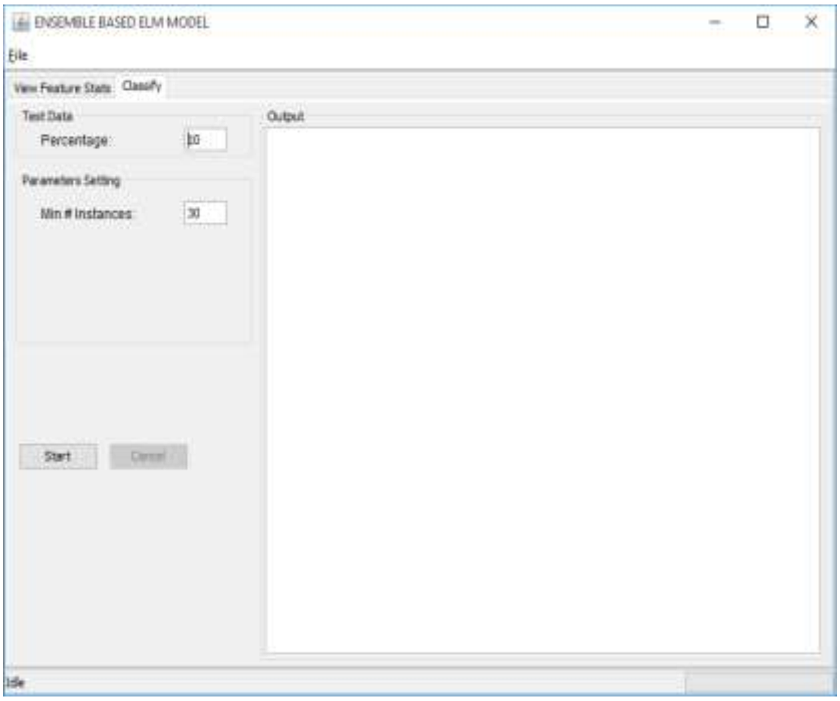

Figure 4: Training and Testing Proposed Model

Using the provided dataset, the proposed model is trained and tested as shown in Figure 4. In this case, 10% of the training data is used as test data for classifying. An evaluation of classification patterns and their accuracy metrics is provided below.

Accuracy rate on the training set: 90.64748201438849 %

Accuracy rate on the test set: 83.33333333333334 %

Time taken: 11.814 s.

Table 2: Performance analysis of classification accuracy over ovarian dataset.

Cancer Dataset	Random Forest+NN	ACO+NN	PSO+NN
Ensemble With Proposed approach(WELM)	96.47	95.75	97.854
Ensemble Without Basic ELM	91.86	89.751	94.981

Table 2 shows how our suggested weighted ELM model improves the ensemble classifier. On a cancer dataset, the true positive and true negative rates are used to calculate the accuracy. It is obvious from the table that our proposed weighted ELM in ensemble classifier improved the classification rate on a medical dataset.

Table 3: Performance analysis of classification accuracy over ovarian dataset.

Diabetes Dataset	Random Forest+NN	ACO+NN	PSO+NN
Ensemble With Proposed approach(WELM)	94.67	96.75	96.92
Ensemble Without Basic ELM	87.24	93.19	92.67

Table 2 shows how our suggested weighted ELM model improves the ensemble classifier. True positive and true negative rates for diabetes diagnoses are used to calculate accuracy. It is obvious from the table that our proposed weighted ELM in ensemble classifier improved the classification rate on a medical dataset.

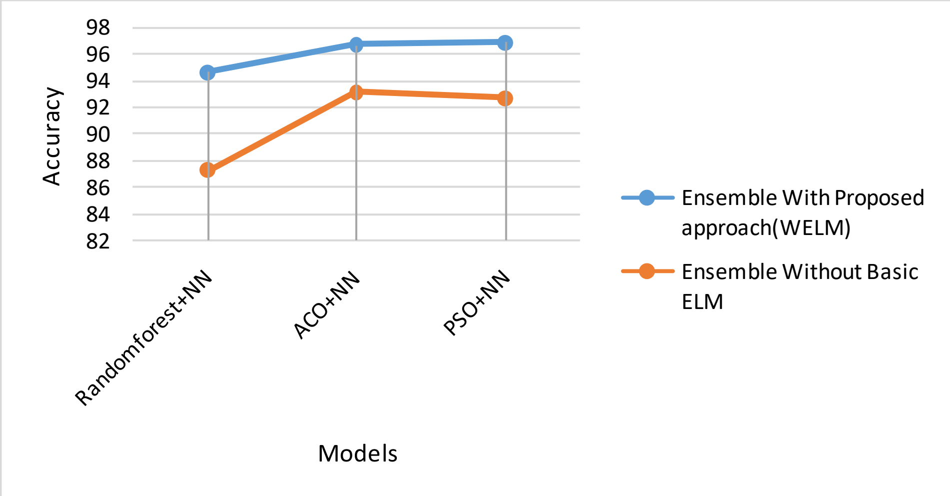

Figure 3: Performance of diabetes disease using classification models.

The graph above shows how the suggested weighted ELM technique improves individual weak classifiers. An ensemble classifier combining the suggested and existing classifiers improved classification rates, as depicted in Figure 1.

Conclusion

This research introduces a novel ensemble classifier weighted extreme learning version for medical data sets. In order to reduce the number of neurons while maintaining the ensemble model's effectiveness, the suggested model employs a fast weighted and error handling approach. When compared to traditional ELM ensemble and simple algorithms, results suggest that weighted ELM in ensemble classification can improve the true positive performance, accuracy, and error rate. Using deep learning models and larger datasets, this research can be applied to pattern pruning on larger datasets in the future.

REFERENCES

[1] M. M. Yusof, R. Mohamed and N. Wahid, " Benchmark of Feature Selection Techniques with Machine Learning Algorithms for Cancer Datasets"," Proceedings of the International Conference on Artificial Intelligence and Robotics and the International Conference on Automation, Control and Robotics Engineering ACM", 2016.

[2] N. Pérez, M. A. Guevara, A. Silva, I. Ramos and J. Loureiro, "Improving the performance of machine learning classifiers for Breast Cancer diagnosis based on feature selection",

"Proceedings of Federated Conference on Computer Science and Information Systems, pp.209–217", 2014.

[3] A. Arafi, R. Fajr and A. Bouroumi, "Breast Cancer Data Analysis Using Support Vector Machines and Particle Swarm Optimization ", "Complex Systems (WCCS), 2nd World Conference on IEEE, pp.1-6, 2014.

[4] Lan Y, Soh YC, Huang GB (2010) Constructive hidden nodes selection of extreme learning machine for regression. Neurocomputing 73(16–18):3191–3199

[5] S. Begum, D. Chakraborty and R. Sarkar, "Identifying cancer biomarkers from leukemia data using feature selection and supervised learning", "IEEE First International Conference on Control, Measurement and Instrumentation", 2016.

[6] Huang G-B, Wang D (2013) Advances in extreme learning machines (ELM2011). Neurocomputing 102:1–2

[7] Zong W, Huang G-B, Chen Y (2013) Weighted extreme learning machine for imbalance learning. Neurocomputing 101:229–242

[8] Huang G-B, Lei C, Siew CK (2006) Universal approximation using incremental constructive feedforward networks with random hidden nodes. IEEE Trans Neural Netw 17(4):879–892

[9] Yu Q, Miche Y, Eirola E, van Heeswijk M, Se´verin E, Lendasse A (2013) Regularized extreme learning machine for regression with missing data. Neurocomputing 102:45–51

[10] Huang G-B, Siew CK (2004) Extreme learning machine: RBF network case. In: International conference of the control, automation, robotics and vision, pp 1029–1036

[11] Cai D, He X, Han J (2008) Training linear discriminant analysis in linear time. In: IEEE 24th international conference on data engineering (ICDE 2008), pp 209–217

[12] Castan˜o A, Ferna´ndez-Navarro F, Herva´s-Martı´nez C (2013) PCA-ELM: a robust and pruned extreme learning machine approach based on principal component analysis. Neural Process Lett 37(3):377–392

[13] http://datam.i2r.a-star.edu.sg/datasets/krbd/index.html